DRAFT

REQUIREMENTS FOR ERNE QA PROCESSES

VERSION: 0.1

AUTHORS: SUZANNA REID & GREG WARNICK

AUGUST, 2007

REVISION CHART

Version	Date	Primary Author	Description
0.1	8/08/2007	Suzanna Reid	Initial Draft

TABLE OF CONTENTS

1.Purpose

This document is intended to describe ERNE QA processes including information regarding responsibilities for the design, build, test, implementation and monitoring of the QA processes. This information will be of value during the initiation of QA workflows and later as a reference document for those involved in maintenance and/or oversight of ERNE. Ongoing quality Assurance is an important component for ERNE due to potential for inaccurate specimen queries because of site product server unavailability at one or more of the numerous sites, inaccurate mapping of data between a site and ERNE hub and query system, alterations in product server data repositories without efficient communication with ERNE developers and/or non-productive query processes.

2.Goals and Success Criteria

The intent is to document QA processes that will assist in monitoring and determining quality of ERNE data. Some processes may be manual while others are automated. Reports generated by each process should allow reviewers to determine how well ERNE infrastructure is functioning and the quality of data or accuracy of representation of data contained at various participating sites.

3.Assumptions and Contraints

The QA processes listed in this document may not be a complete list but is instead, a work in progress. Quality Assurance will be phased in as needs are identified.

4.Definitions and Acronyms

ERNE – EDRN Resource Network Exchange.

DMCC – Data Management Coordination Center.

JPL – Jet Propulsion Laboratory

Site – Each labe/site/location/University/etc participating in providing data to ERNE.

Product Servers – servers set up at each site to participate in ERNE queries. These servers may contain only a subset of actual study data that satisfies the needs for ERNE queries.

5.References

ERNE project information, in general, can be found at one of the following links listed below:

DMCC:
N:\edrn\informatics\ernephase4 (note: only accessible to DMCC employees)

Examples of previous QA comparing ERNE queries with actual site queries:
DMCC examples of QA processes-1: N:\edrn\informatics\qa\dec2004
DMCC examples of QA processes-2: N:\edrn\informatics\qa\feb2003

Other QA documents:
N:\edrn\informatics\erne_qa_process.ppt (Note: as of 08/2007 this presentation may be out of date but still provides some information on original QA process design)

Spreadsheet first used in 08/2007 to capture QA processes and high level design:
N:\edrn\informatics\ernephase4\qaplan.xls

JPL:
Wiki site: http://oodt.jpl.nasa.gov/wiki/display/edrn/Home
OODT Wiki site: http://oodt.jpl.nasa.gov/wiki/display/oodt/Home

EDRN:
RN's ERNE site: https://ginger.fhcrc.org/edrn/imp/GateServlet_2

6.ERNE QA TEAM INFROMATION.

DMCC: Greg Warnick and Suzanna Reid
JPL: Sean Kelley
Site: (list sites and contacts here)

7.FUNCTIONAL REQUIREMENTS (FEATURES)

7.1Statistics of Each Site's ERNE Product Server Up Time

7.1.1 General Description of Product Server Availability.

Product servers contributing data to the ERNE **cache** will only need to be up at the time JPL code runs to retrieve data into the cache. Therefore, the availability of site product servers at the time of caching will be tracked and reported. On the other hand, product servers at sites not contributing to cached data will be pinged frequently and tracked for reporting purposes. For both types of product servers an email will be sent when server(s) are found to be down. This will serve as an alert to the IT support team for follow up, root cause analysis and subsequent actions to bring server(s) back in line with ERNE.

7.1.2 Target Technology of QA Process.

Product Servers at each site.

7.1.3 Method of QA Process

7.1.3.1 For Site Product Servers Participating in ERNE Caching.

7.1.3.1a Ping server immediately prior to running the job that extracts data into the cache.

7.1.3.1b Write date, time and up/down status to database (create database or table to use for tracking server metrics)

7.1.3.1c If server is down and unavailable (resulting in failed attempt to refresh cached data for that site), send email to DMCC, JPL and site regarding server unavailability

7.1.3.2 For Site Product Servers **not** participating in ERNE Caching.

7.1.3.2a Ping servers hourly and record date, time and up/down status to database.

7.1.3.2b If server is down and unavailable for searching, send email to DMCC, JPL and site regarding server unavailability.

7.1.4 Accountability for QA Process

7.1.4.1 **Design**: JPL

7.1.4.2 **Build**: JPL

7.1.4.3 **Test**: JPL

7.1.4.4 **Implement**: JPL

7.1.4.5 **Ongoing monitoring**: JPL, DMCC and Sites

7.1.5 Metrics Reporting

7.1.5.1 Contents of Metrics Report

7.1.5.1a **For sites participating in caching:** % Successful cache attempts for given reporting period listed per site (how often was a site's product server and data available for nightly caching?).

7.1.5.1b **For sties not participating in caching:** Report % of server up time for given reporting period listed per site based on hourly ping of server.

7.1.5.2 How Often Report is to be generated.

Report will be generated and distributed on a quarterly basis.

7.1.5.3 Audience for Report

JPL, DMCC, NCI and Sites

7.1.6 Response to QA Report Issues Identified.

Any site with site unavailability below a threshold of tolerance (TBD) will work with JPL and DMCC if necessary to resolve availability issue. DMCC will facilitate coordination of efforts to resolve product server issues when need is identified from quarterly report.

7.2Comparison of ERNE Search Results Based on Mapping (non-cached data) Against Site's Results Based on Direct Query without Mapping of Product Servers.

7.2.1 General Description of QA purpose.

Each site provided "mapping" information to the ERNE project team attempting to match fields in the databases maintained on each site's product server with the intended query targets of ERNE. Thus, as queries are run through ERNE the results are representative of "mapped" or interpreted data. This QA function is to determine if ERNE mapping assumptions continue to be correct and match each site's actual data without mapping interpretations. Canned queries will be run based on mapping as well as queries written by each site directly against databases (non-mapped). Results will be compared and tracked for any drift patterns.

7.2.2 Target Technology of QA Process.

Site's product server.

7.2.3 Method of QA Process

7.2.3.1 Develop queries (simple and complex) based on each site's mapping instructions. **Note:** Queries may change if QA Process #4 changes.

7.2.3.2 Have each site build (or assist each site in building) queries to match search criteria developed for mapped data. The queries built by each site should not take into consideration any previous mapping work or documentation.

7.2.3.3 Run both sets of queries again each site's product server

7.2.3.4 Automate comparison of search results to examine variances.

7.2.3.5 If variances differ by >1%, email will be sent to DMCC, JPL and the site.

7.2.4 Accountability for QA Process

7.2.4.1 **Design:** DMCC , JPL and individual sites.

7.2.4.2 **Build:** JPL

7.2.4.3 **Test:** JPL

7.2.4.4 **Implement:** JPL

7.2.4.5 **Ongoing Monitoring of QA Process (list QA process here).** JPL and DMCC

7.2.5 Reporting Mechanism

7.2.5.1 How Often Report is to be generated.

Queries and results comparison will be run monthly. Metrics on variances will be captured, stored and reviewed by the DMCC and JPL. Variances by month (for year to date) will be charted and reported quarterly.

7.2.5.2 Audience for Report

DMCC, JPL and Sites

7.2.6 Response to QA Report Issues Identified.

If variances > 1% persist, a mapping review should be completed.

7.3Comparison of ERNE Cached Data Query Results Against a Site's Product Server (a.k.a live) Data Query Results.

7.3.1 General QA Task Description.

Queries will be built (can be similar to those built for QA #2) and run against the cached data as well as each product server directly. The query results will be compared and metrics should indicate discrepancies between cached and non-cached data.

7.3.2 Target Technology of QA Process.

Cached data and product servers.

7.3.3 Method of QA Process

7.3.3.1 Build matching queries against cached data and mapped product server data.

Note: These queries can be the same or similar to those built for QA process #2.

7.3.3.2 Run, in parallel, matching queries against ERNE cache and product servers weekly.

7.3.3.3 Run automated comparison of search results capturing variances based on sites and fields queried.

7.3.3.4 DMCC will review variance data looking for trends.

7.3.4 Accountability for QA Process

7.3.4.1 **Design:** JPL

7.3.4.2 **Build:** JPL

7.3.4.3 **Test:** JPL

7.3.4.4 **Implement:** JPL

7.3.4.5 **Ongoing Monitoring of QA Process (list QA process here).:** JPL

7.3.5 Reporting Mechanism

7.3.5.1 How Often Report is to be generated.

Queries will be run and variance reports generated on a weekly basis. Variances will be reported for each site and each field queried. A report reviewing average variances across all fields for a site, per month and year to date will also be generated.

7.3.5.2 Audience for Report

DMCC and JPL (perhaps NCI on a quarterly basis for average variance data).

7.3.6 Response to QA Report Issues Identified.

If DMCC identifies trends where variances between cached data and a site's data are persistent or intensify, root cause analysis will be coordinated by the DMCC and involve JPL and sites if necessary. Based on evidence from the root cause analysis, action will be taken to resynchronize cached data to product server data.

7.4Comparison of a site's Original Mapping Against the Possible Changes to The Site's Database (either data or structure).

7.4.1 General QA Task Description.

This QA process basically involves a mapping review and audit. JPL, DMCC and each site will participate in a review of changes to data captured and reported on the product servers. For instance, it is possible that new data (a new column) is added to the product server data base requiring additional mapping or inclusion in ERNE query criteria and cache. Or, an existing data field (column) on the product server may change. Examples would include a format changes or the type of data collected (inclusion of additional permissible values). These changes should always be communicated to JPL and DMCC but the audit should help if communications were missed or insufficient.

7.4.2 Target Technology of QA Process.

Product servers and ERNE cache.

7.4.3 Method of QA Process

7.4.3.1 Queries will be run to determine if any additional data, in terms of permissible values, is captured by any of the sites data fields on the product servers. Are new or different values being reported in the data fields?

7.4.3.2 Query results from 7.4.3.1 will be reported to each site along with a checklist of items to be reviewed regarding product server data. Discussion points and items will include:

a) Have there been any changes to the underlying data?

b) Have any new data fields been added?

c) Have any previously mapped but unreported data fields been added?

d) Has the data captured in any of the existing data fields changed (including addition of permissible values)?

7.4.4 Accountability for QA Process

7.4.4.1 **Design:** JPL & DMCC

7.4.4.2 **Build:** JPL & DMCC

7.4.4.3 **Test:** JPL

7.4.4.4 **Implement:** JPL & DMCC

7.4.4.5 **Ongoing Monitoring of QA Process (list QA process here).:** JPL & DMCC

7.4.5 Reporting Mechanism

7.4.5.1 How Often Report is to be generated.

Audit and report will be done once per year per site.

7.4.5.2 Audience for Report

DMCC, JPL and site

.6 Response to QA Report Issues Identified.

ERNE cache and queries will be altered if necessary based upon discoveries in product server data changes.

7.5Daily Query Status Report

7.5.1 General QA Task Description.

Verify servers are not only up but can return productive ERNE query results. Two searches will be run daily (number of participants and number of specimens) known to return data from participating sites. QA Process #1 pings each server to determine availability whereas this QA process confirms that search capabilities are intact.

7.5.2 Target Technology of QA Process.

ERNE query handler and product servers.

7.5.3 Method of QA Process

7.5.3.1 Two queries, one for number of participants and one for number of specimens, will be written for each site.

7.5.3.2 Queries will run against production servers at each site on a daily basis.

7.5.3.3 Query results, per site, will be logged to a database.

7.5.3.4 If any query results are returned with zero value, JPL, DMCC and the site returning a zero result will receive email alerts specifying the queries and results.

7.5.4 Accountability for QA Process

7.5.4.1 **Design:** DMCC and JPL

7.5.4.2 **Build:** JPL

7.5.4.3 **Test:** JPL

7.5.4.4 **Implement:** JPL

7.5.4.5 **Ongoing Monitoring of QA Process (list QA process here):** JPL

7.5.5 Reporting Mechanism

7.5.5.1 How Often Report is to be generated.

Email alerts will be sent as needed when query result(s) are zero for any given site. Metric results indicating percentages of time product servers are available for **query** at each site can be reported to a broader audience (NCI) in a quarterly or yearly report. Query result log can be used to create additional reports.

7.5.5.2 Audience for Report

Email alerts will be sent to DMCC, JPL and Site returning zero search results. Quarterly and/or yearly metrics can be included with other reporting metrics and shared with NCI.

.6 Response to QA Report Issues Identified.

Once email alert is received, DMCC will coordinate root cause analysis with both JPL and the implicated site's participation. Short term and long term changes will be implemented based upon root cause analysis.

7.6Contact verification

7.6.1 General QA Task Description.

Confirm each site's contact person's name, e-mail and telephone number for collaborative investigator specimen requests as well as for ERNE technical support and ERNE database mapping. The DMCC will request information updates from each site on a yearly basis.

7.6.2 Target Technology of QA Process.

NA

7.6.3 Method of QA Process

The DMCC will email sites for an updated status on contact names, email addresses, phone numbers, fax numbers, back up contacts, etc for ERNE related work and issues. The DMCC will coordinate these efforts via email and phone contact.

7.6.4 Accountability for QA Process

7.6.4.1 **Design:** DMCC

7.6.4.2 **Build:** NA

7.6.4.3 **Test:** NA

7.6.4.4 **Implement:** DMCC

7.6.4.5 **Ongoing Monitoring of QA Process (list QA process here):** DMCC

7.6.5 Reporting Mechanism

7.6.5.1 How Often Report is to be generated.

Once a year

7.6.5.2 Audience for Report

Updates to contact list will be shared with DMCC staff and JPL.

.6 Response to QA Report Issues Identified.

NA

7.7User statistics and monitoring.

7.7.1 General QA Task Description.

This QA process will help identify ERNE issues with unproductive searches due to timeouts and/or returned errors. This QA process will also provide metrics regarding the number of searches and the types of searches.

7.7.2 Target Technology of QA Process.

ERNE Query handler.

7.7.3 Method of QA Process

7.6.3.1 For every search, log the following information:

a) The search performed (log parameters searched or query string).

b) Were results returned (Y/N)?

c) Was there a timeout issue?

d) Was there an error returned?

7.7.4 Accountability for QA Process

7.6.4.1 **Design:** JPL

7.6.4.2 **Build:** JPL

7.6.4.3 **Test:** JPL

7.6.4.4 **Implement:** JPL

7.6.4.5 **Ongoing Monitoring of QA Process (list QA process here)**: JPL

7.7.5 Reporting Mechanism

7.7.5.1 How Often Report is to be generated.

A report containing the following metrics should be reported on a monthly basis: number of searches; percentage of successful searches, number of time outs, percentage of searches resulting in time outs, number of search errors, percentage of searches returned with errors, most commonly searched parameters, other?

7.7.5.2 Audience for Report.

Report will be shared with DMCC, JPL, Sites and NCI (perhaps reported to NCI only on a quarterly basis).

7.7.6 Response to QA Report Issues Identified.

If report metrics indicate issues with intolerable percentages of timeouts and errors, root cause analysis will be coordinated by the DMCC and action taken to correct issues based on findings. Search parameters will be used to analyze additional search criteria that might facilitate the user's searches.

7.8End User Validation of Search Results.

7.8.1 General QA Task Description.

A Quality Control individual will be identified at the DMCC and periodically run searches using the ERNE user interface for all participating sites. The Quality Assurance individual will review the returned search results and identify glaring errors and suspicious data. This is a subjective test requiring no automation. This process is necessary because automated processes cannot subjectively scan for search issues.

7.8.2 Target Technology of QA Process.

NA

.3 Method of QA Process

.3.1 DMCC Quality Control person will run an advanced search on each participating ERNE site. Multiple search combinations should be for each site and can be determined by individual performing QC.

.3.2 The search results should be reviewed and suspicious or illogical data indicated. The following gives an example of the types review to perform:

a) Do the dates make any sense?
b) Do the number of participants and samples returned make sense.
c) Does the participants ages seem reasonable?
d) Can text in report fields be read and interpreted?

7.8.3.3 Identification of suspicious data should trigger DMCC to initiate coordination of root cause analysis with JPL and sites participating as needed.

7.8.4 Accountability for QA Process

7.6.4.1 **Design:** DMCC

7.6.4.2 **Build:** NA

7.6.4.3 **Test:** NA

7.6.4.4 **Implement:** DMCC

7.6.4.5 **Ongoing Monitoring of QA Process (list QA process here)**: NA

7.8.5 Reporting Mechanism

7.6.5.1 How Often Report is to be generated.

This QA process should be performed on a quarterly basis.

7.6.5.2 Audience for Report.

NA

.6 Response to QA Report Issues Identified.

DMCC will flag and communicate issues as needed and coordinate research and response to identified problems.

8.Ranking of Functional Requirements

Ranking	ERNE QA Process #	QA Process title
1	QA8 : requirement 7.8	**End User Validation of Search Results.**
2	QA1 : requirement 7.1	Statistics of Each Site's ERNE Product Server Up Time.
3	QA5 : requirement 7.5	Daily Query Status Report
4	QA7 : requirement 7.7	User statistics and monitoring
5	QA2 : requirement 7.2	Comparison of ERNE Search Results Based on Mapping (non-cached data?) Against Site's Results Based on Direct Query without Mapping of Product Servers
6	QA3 : requirement 7.3	Comparison of ERNE Cached Data Query Results Against a Site's Product Server (a.k.a live) Data Query Results.
7	QA4 : requirement 7.4	Comparison of a site's Original Mapping Against the Possible Changes to The Site's Database (either data or structure).
8	QA6 : requirement 7.6	User statistics and monitoring

.

9.Usability

This section should provide information about the characteristics relating to the effort needed for use by a stated or implied set of users. This may include:

- Understandability
- Learn ability
- Operability

The presentation style in this section is at project manager's discretion.

10.Platform Dependencies (Hardware and Software)

Explain what platforms are expected (server and client). Explain network configuration requirements. These may be detailed in a design document which can be referenced here.

Explain what platforms the client presentation layer must run on (e.g., browser versions). Refer to an installation and or deployment document if applicable.

11. CROSS SYSTEM INTERFACE REQUIREMENTS

This section should provide information about the characteristics relating to software's ability to interact with other systems. It may include requirements about converting data to import into the system. Detail the mapping needed from the old system to the new system, if applicable.

DRAFT